Metabolic syndrome in Xinjiang Kazakhs and construction of a risk prediction model for cardiovascular disease risk

Lei Mao[1]☯, Jia He[1]☯, Xiang Gao[1,2], Heng Guo[1], Kui Wang[1], Xianghui Zhang[1], Wenwen Yang[1], Jingyu Zhang[1], Shugang Li[1], Yunhua Hu[1], Lati Mu[1], Yizhong Yan[1], Jiaolong Ma[1], Yusong Ding[1], Mei Zhang[1], Jiaming Liu[1], Rulin Ma[1]*, Shuxia Guo[1,3]*

1 Department of Public Health, Shihezi University School of Medicine, Shihezi, Xinjiang, China,
2 Department of Nutritional Sciences, The Pennsylvania State University, University Park, United States of America, 3 Department of Pathology and Key Laboratory of Xinjiang Endemic and Ethnic Diseases (Ministry of Education), Shihezi University School of Medicine, Shihezi, Xinjiang, China

☯ These authors contributed equally to this work.
* marulin@126.com (RM); gsxshzu@sina.com (SG)

Abstract

Background

The high prevalence of metabolic syndrome (MetS) and cardiovascular diseases (CVD) is observed among Kazakhs in Xinjiang. Because MetS may significantly predict the occurrence of CVD, the inclusion of CVD-related indicators in metabolic network may improve the predictive ability for a CVD-risk model for Kazakhs in Xinjiang.

Methods

The study included 2,644 subjects who were followed for 5 years or longer. CVD cases were identified via medical records of the local hospitals from April 2016 to August 2017. Factor analysis was performed in 706 subjects (267 men and 439 women) with MetS to extract CVD-related potential factors from 18 biomarkers tested in a routine health check-up, served as a synthetic predictor (SP). We evaluated the predictive ability of the CVD-risk model using age and SP, logistic regression discrimination for internal validation (n = 384; men = 164, women = 220) and external validation (n = 219; men = 89, women = 130), calculated the probability of CVD for each participant, and receiver operating characteristic curves.

Results

According to the diagnostic criteria of JIS, the prevalence of MetS in Kazakh was 30.9%. Seven potential factors with a similar pattern were obtained from men and women and comprised the CVD predictors. When predicting CVD in the internal validation, the area under the curve (AUC) were 0.857 (95%CI 0.807–0.898) for men and 0.852 (95%CI 0.809–0.889) for women, respectively. In the external validation, the AUC to predict CVD were 0.914 (95%CI 0.832–0.963) for men and 0.848 (95%CI 0.774–0.905) for women. It is suggested

that SP might serve as a useful tool in identifying CVD with in Kazakhs, especially for Kazakhs men.

Conclusions

Among **7** potential factors were extracted from 18 biomarkrs in Kazakhs with MetS, and SP may be used for CVD risk assessment.

Introduction

A recent statistical report [1] indicates that China has 290 million people with cardiovascular disease (CVD), with a mortality rate higher than those for cancer and other diseases, accounting for more than 40% of all deaths. Moreover, the CVD mortality rate was higher in rural areas than in urban areas. The increased risk of CVD was due to the accumulation of risk factors such as obesity, dyslipidemia, hyperglycemia, and high blood pressure, and these risk factors are the major components of metabolic syndrome (MetS). Although the definition varies [2–6], MetS includes 4 primary components: hypertension, hyperlipidemia, central obesity, and impaired glucose tolerance. Each component interacts with other components to form a complex metabolic network [6]. The MetS metabolic network also consider serum liver enzyme levels [7,8], liver function index [9,10], and renal function index [11,12]. However, it remains unclear whether these indicators can be used as predictors in a CVD-risk model.

Many CVD risk assessment tools have been described, the most commonly cited one is the Framingham Study (FRS) model. Because of the different research populations, FSR could not be applied to all groups[13]. Subsequently, many studies have increased the predictive power of FRS for CVD by refining the study populations [14], increasing the number of risk factors [15], and using appropriate predictors [16]. Moreover, the CVD predictive model is constructed using the Whitehall integral model [17], neural network model [18], QRISK score [19], and components of MetS extracted by exploratory factor analysis (EFA) as CVD predictors [20]. These predictive models analyzed all aspects of CVD, and continuously supplemented the shortcomings of FRS prediction. However, the above CVD predictive models have local advantages and weak extrapolation ability [14]. Therefore, in order to effectively prevent and treat CVD, it is necessary to construct a targeted CVD prediction model.

Ethnic Kazakh is the most populous nomadic ethnic group in Xinjiang, China, and is mainly resides in mountain pasturelands in northern Xinjiang. Previous research found that the prevalence of MetS, hypertension, and obesity were higher in Kazakhs than in other ethnic groups in the same area [21–23]. Therefore, this study used EFA to identify potential factors associated with MetS in Kazakh patients. Latent factors have been used to build a predictive model of CVD risk for the Kazakh population in rural Xinjiang. The results provided the basis for the developing interventions for CVD and other chronic diseases in Xinjiang Kazakhs and may be of important reference value for the prediction and control of CVD for the Xinjiang area extending to the Belt and Road countries including Kazakhstan.

Materials and methods

Study population

The Kazakh residents were from 6 villages in Xinyuan County, Xinjiang Kazakh Autonomous Region. The baseline survey was conducted from April 2009 to December 2013, and was

completed in 2,644 permanent residents older than 18 years (1,085 men and 1,559 women). Between April 2016 and August 2017, 2,286 subjects from the baseline population who participated for 5 years or more in the follow-up survey were evaluated for the incidence of CVD; the follow-up rate was 86.46%. Of the them in the baseline population, 706 were first diagnosed with MetS according to the Joint Interim Statement [6], with a prevalence of 30.88% (the age-adjusted prevalence of 28.61%), including 267 men and 439 women. Exclusion criteria at baseline identified 281 subjects with CVD and the remaining 2005 were analyzed. Of 706 subjects with MetS, 191 (82 men and 109 women) had CVD diagnosed by physicians, as case group. Controls (n = 384; men = 164, women = 220) were randomly selected from individuals without MetS or CVD in the case-control study design. These subjects were used for internal verification. Another Kazakh research site, Halabula Township, was the source of 243 external verification controls, selected in August 2017 at the end of the information collection period. Of these, 24 were excluded because of CVD or lack of information at baseline, and 219 (89 men, 130 women) were included for analysis. By the end of the study, CVD developed in 30 of the 219. The cumulative incidence rate was 13.7% (12.4% for men and 14.6% for women).

Selection and measurement of biomarkers

We extracted 18 kinds of CVD related routine examination index from the MetS metabolic network, including body weight, waist circumference (WC), body adiposity index (BAI), systolic blood pressure (SBP), diastolic blood pressure (DBP), high-density lipoprotein cholesterol (HDL-C), apolipoprotein A (APOA), fasting blood glucose (FBG), fructosamine (FMN), alanine aminotransferase (ALT), aspartate aminotransferase (AST), α-hydroxybutyrate dehydrogenase (α-HBDH), total bilirubin (TBIL), indirect bilirubin (IBIL), albumin (ALB), uric acid (UA), serum creatinine (CREA) and blood urea nitrogen (BUN). Among them, WC, SBP, DBP, HDL-C and FBG were all defined in the traditional MetS. Also, others have been confirmed in the study to meet the above criteria[7–12]. All blood samples were analyzed using an automatic biochemical analyzer (Olympus AU 2700; Olympus Diagnostics, Hamburg, Germany).

Methods

All participants filled out questionnaires covering demographic data and the history of disease. Physical examination including height, weight, WC, hip circumference and blood pressure, height and weight requirements of wearing light clothes to measure. BAI was calculated as [hip circumference (cm) / (height (m))$^{1.5}$–18]. After each subject was rested for 15 min, each subject measured the blood pressure three times using a mercury sphygmomanometer and then calculated the average value. Each participant signed an informed consent form. The study was approved by the Institutional Ethics Review (IERB) for the First Affiliated Hospital of Shihezi University (IERB No.SHZ2010LL01). Outcome information was obtained directly by examining the medical records of the local hospital discharge records and medical insurance.

Definition of MetS and CVD

Based on the diagnostic criteria recommended by the JIS[6], MetS was defined as presence of three or more of the following five risk factors: 1) WC≥85 and 80 cm for male and female, respectively; 2) Triglyceride (TG)≥1.70mmol/L; 3) HDL-C<1.00 mmol/L in male and HDL-C<1.30 mmol/L in female; 4) SBP≥130mmHg or DBP≥85mmHg; 5) FPG≥5.6mmol/L.

The following event is identified as an ending event: 1) Hospitalizations for CVD at follow-up; coronary intervention (cardiac catheterization or coronary bypass), angina (or nitroglycerin after cohort study) and CVD death (ICD9: codes 390–495); 2) According to the hospital

medical records of hospitalized dued to the following reasons: coronary artery atherosclerosis, coronary heart disease, unstable angina, myocardial infarction, heart failure, stroke, transient cerebral ischemia and peripheral vascular disease (abdominal aortic aneurysm, peripheral vascular surgery or carotid endarterectomy); 3) The CVD events were recorded according to the hospitalization records and questionnaires. If two or more events of the same class of subjects occur, the first occurrence was the outcome event.

Statistical analysis

Descriptive analysis. For patients with MetS, Student's t test (for continuous variables) was used to evaluate significant differences between men and women for 18 biomarkers.

Steps of the development of synthetic predictor. EFA with principal component algorithm and varimax rotation from correlation matrix was performed to extract independent factors of MetS from above 18 manifest biomarkers for male and female MetS groups respectively. The criteria for retaining factors were set up as eigenvalue>1 as well as accounting for 75% of the total variation. Only variables that shared at least 15% of the factor variance, corresponding to a factor loading of at least 0.45 were used for further analytical interpretation. After EFA, the clinical significance of each latent factor was named, and Synthetic predictor (SP) [19] was created using a weighted approach: $SP = \gamma_1 F_1 + \gamma_2 F_2 + \ldots + \gamma_n F_n$, where F_1, F_2, $\ldots F_n$ were the extracted independent factors with specific clinical significance from the 18 manifest biomarkers, and γ_1, γ_2, $\ldots \gamma_n$ denoted their risks to CVD, which were partial regression coefficients in LRD regression models described below.

$$P = \frac{\exp(B)}{1 - \exp(B)} \qquad (1)$$

The logistic regression discrimination (LRD), on the basis of the extreme case-control design, was performed using the model, where P was the probability of CVD. B denoted the discriminant vector estimated by logistic regression, where $B = \beta_0 + \beta_1 age + \gamma_1 F_1 + \gamma_2 F_2 \cdots \gamma_n F_n$ or $B = \beta_0 + \beta_1 age + SP$.

According to the logistic regression discrimination, the results of the internal verification population and the external verifying crowd were evaluated respectively. The probability of each participants with CVD was calculated and the ROC curve was analyzed with the actual CVD follow-up outcome. Calculated the sensitivity and specificity of the disease probability at each cut-off point, and determine the cut-off point with superior sensitivity and specificity.

All statistical analysis for the predictive models was performed using SPSS version 19.0 for Windows. The area under the curve (AUC) for the receiver operating characteristic (ROC) curve analysis, together with the sensitivity, specificity, and cutoff P values, were determined using MedCalc software. A two-sided P value <0.05 was considered statistically significant.

Results

In our study, the prevalence of MetS was 30.88% (the standardized prevalence rate was 28.61%) at baseline. 359 CVD events were diagnosed, with a prevalence of 13.87%. Table 1 shows the distribution of age and 18 biomarkers between men and women with MetS. All variables were significantly different between genders. Of these biomarkers, age, weight, WC, SBP, DBP, FBG, FMN, ALT, AST, TBIL, IBIL, ALB, UA, CREA and BUN were higher in men than in women, while BAI, HDL-C, APOA and α-HBDH were higher in women than in men. S1 Table showed the baseline population distribution of age and the eighteen biomarkers between male and female groups, indicating that all variables except FMN, α-HBDH and MS were significantly different between men and women.

Table 1. Distribution of age and the eighteen biomarkers between male and female metabolic syndrome groups.

	Male(n = 267)		Female(n = 439)		P value
	mean	SD	mean	SD	
Age (years)	50.26	12.80	48.29	12.67	0.046
Weight (kg)	75.72	12.40	65.70	12.32	P<0.001
Waistline (cm)	94.20	10.59	89.87	10.74	P<0.001
BAI	27.65	3.63	32.19	4.78	P<0.001
SBP (mmHg)	144.94	21.43	140.77	26.38	0.030
DBP (mmHg)	92.43	13.66	89.08	15.33	0.004
HDL-C (mmol/L)	1.20	0.41	1.26	0.38	0.042
APOA (g/l)	1.21	0.30	1.24	0.28	0.030
FBG (mmol/L)	6.07	2.14	5.66	1.40	0.002
FMN (umol/l)	228.36	63.79	215.26	36.14	0.001
ALT (IU/L)	21.30	15.27	16.88	13.22	P<0.001
AST (IU/L)	33.02	27.56	26.78	15.88	P<0.001
α-HBDH (IU/L)	127.52	40.73	134.06	40.45	0.040
TBIL (umol/l)	11.28	6.70	9.28	5.26	P<0.001
IBIL (umol/l)	8.55	5.29	6.88	4.09	P<0.001
ALB (g/l)	41.72	8.62	39.13	8.84	P<0.001
UA (umol/L)	288.77	98.23	213.00	75.64	P<0.001
CREA (umol/l)	70.43	13.43	56.93	12.27	P<0.001
BUN (mmol/l)	4.89	1.40	4.35	1.21	P<0.001

Note: BAI: Body adiposity index; SBP: Systolic blood pressure; DBP:Diastolic blood pressure; HDL-C: High-density lipoprotein cholesterol; APOA: Apolipoprotein A; FBG: Fasting blood-glucose; FMN: Fructosamine; ALT: Alanine aminotransferase; AST: Aspartate transferase; α-HBDH: α-Hydroxybutyrate dehydrogenase; TBIL: Total bilirubin; IBIL:Indirect bilirubin; ALB: Serum albumin; UA: Serum uric acid; CREA: Creatinine; BUN: Blood urea nitrogen.

S2 Table shows significant correlations for most biomarkers. In all subjects, weight, WC, SBP, DBP, HDL-C, APOA, ALT, AST, α-HBDH, TBIL, IBIL, ALB, UA, and CREA were significantly correlated with most variables analyzed. The BAI, FBG, FMN and CREA were correlated with approximately half of the variables analyzed. The Kaiser-Meyer-Olkin test values in the male and female groups were 0.636 and 0.632 respectively, and Bartlett's sphericity test for 2 groups had a significance of P<0.001. This suggested that the variables were not independent and were strongly correlated, making them suitable for EFA.

Table 2 shows the explained variance, cumulative variance, and loading of the first 7 factors. The results suggested that total variance of 74.02% and 73.74% was explained by the first 7 factors for men and women, respectively. Following the criteria of interpretation mentioned in the statistical analysis section, 7 independent factors with their specific clinical significance were retained and named for 2 groups respectively. For the male group, the first factor was named Obesity factor (OF) and contributed by weight & WC & BAI, following Hepatic function factor (HFF) by TBIL & IBIL, Lipid factor (LF) by HDL-C & APOA & ALB, Renal metabolic factor (RMF) by UA & CREA & BUN, Enzyme metabolism factor (EMF) by ALT & AST & α-HBDH, Blood pressure factor (BPF) by SBP & DBP, and Glucose metabolism factor (GMF) by FBG & FMN. For the female group, the first was named HFF and contributed by TBIL & IBIL & ALB, following OF by weight & WC & BAI, LF by HDL-C & APOA, BPF by SBP & DBP, EMF by ALT & AST & α-HBDH, RMF by UA & CREA & BUN, and GMF by FBG & FMN. Although the order of decline in interpretation of differences between men and women is slightly different, their factor patterns were similar. S3 Table shows the standardized scoring coefficients of each factor for male and female groups.

Table 2. Factor loadings by principal component analysis with varimax rotation on 18 routine health check-up biomarkers in MetS patients.

	Male (n = 267)							Female (n = 439)						
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
	OF	HFF	LF	RMF	EMF	BPF	GMF	HFF	OF	LF	BPF	EMF	RMF	GMF
Weight	**0.867**	0.039	-0.099	0.073	0.032	0.060	-0.021	0.071	**0.860**	-0.115	-0.011	0.100	0.010	-0.002
Waistline	**0.918**	0.006	-0.022	0.019	0.077	0.081	0.046	0.054	**0.910**	-0.010	0.053	0.015	-0.021	0.042
BAI	**0.745**	0.025	0.133	-0.011	-0.089	0.165	0.133	-0.055	**0.750**	0.121	0.222	-0.038	0.060	0.007
SBP	0.131	0.021	0.091	0.028	-0.070	**0.894**	0.078	0.039	0.109	0.054	**0.929**	-0.010	0.002	0.044
DBP	0.148	-0.019	0.049	-0.028	0.061	**0.900**	-0.047	0.094	0.118	-0.006	**0.916**	0.050	0.027	-0.003
HDL-C	0.015	0.103	**0.938**	-0.008	0.053	0.073	-0.003	0.287	-0.012	**0.899**	0.022	0.078	-0.022	0.017
APOA	-0.055	0.077	**0.926**	0.051	0.089	0.110	0.011	0.109	-0.018	**0.906**	0.012	-0.003	0.078	0.075
FBG	0.115	-0.041	0.026	-0.166	0.055	0.046	**0.806**	-0.008	0.061	0.006	0.016	-0.023	0.048	**0.920**
FMN	0.014	0.206	0.001	0.360	0.155	-0.027	**0.596**	0.195	-0.040	0.199	0.036	0.441	0.045	**0.649**
ALT	0.149	0.067	0.014	-0.011	**0.806**	-0.090	0.085	0.117	0.052	-0.004	-0.063	**0.749**	-0.014	0.113
AST	0.033	0.184	-0.008	-0.015	**0.801**	-0.005	0.019	0.266	0.051	-0.064	-0.037	**0.732**	-0.055	0.093
α-HBDH	-0.170	-0.011	0.170	0.068	**0.569**	0.085	0.067	-0.117	-0.021	0.229	0.207	**0.621**	0.185	-0.088
TBIL	-0.013	**0.957**	0.101	0.066	0.113	0.032	0.053	**0.918**	0.005	0.119	0.056	0.171	0.019	0.038
IBIL	0.030	**0.957**	0.077	0.002	0.083	0.003	0.024	**0.912**	-0.002	0.187	0.045	0.033	0.024	-0.007
ALB	0.178	0.440	**0.468**	0.366	0.172	-0.139	0.166	**0.537**	0.096	0.432	0.092	0.203	0.041	0.150
UA	0.187	0.384	0.274	**0.525**	0.299	-0.056	-0.064	0.408	0.302	0.020	0.013	0.120	**0.530**	0.184
CREA	-0.026	0.118	-0.149	**0.806**	0.041	0.022	-0.153	-0.009	-0.002	-0.127	-0.035	-0.078	**0.853**	-0.013
BUN	0.029	-0.122	0.172	**0.786**	-0.112	0.013	0.176	-0.011	-0.039	0.187	0.060	0.093	**0.729**	0.028
% Variance explained	20.450	13.677	10.400	9.000	7.764	7.045	5.687	21.316	13.317	9.397	8.501	8.198	7.040	5.972
Cumulative variance	20.450	34.128	44.528	53.528	61.292	68.338	**74.025**	21.316	34.633	44.030	52.532	60.730	67.770	**73.742**

Note: Factors were named as Obesity factor(OF),Hepatic function factor (HFF), Lipid factor (LF), Enzyme metabolic factor(EMF),Blood pressure factor (BPF), Renal metabolic factor(RMF), Glucose metabolism factor(GMF). **Bold** indicates that the absolute value of the factor loading was >0.45.

Figs 1 and 2 show the AUC with sensitivity, specificity, and the cut-off points of *P* values (criterion) for the LRD model from the internal validation design using age and SP as discriminant factors (for men: B = 0.733–0.039Age + SP; for women: B = 0.077–0.016Age + SP). In males, the AUC was 0.857 (95% confidence interval [CI]: 0.807–0.898), with a Youden's index value of 0.622 (Sen = 80.49%, Spe = 81.71%), and an optimal cut-off of 0.30. In females, the AUC was 0.852 (95% CI: 0.809–0.889), with a Youden's index value of 0.544 (Sen = 88.07%, Spe = 66.36%), and an optimal cut-off of 0.23. Figs 3 and 4 show the AUC with sensitivity, specificity, and cut-off points of *P* value (criterion) by LRD model from the external validation design using age and SP as discriminant factors (for men: B = -7.55 + 0.079Age + SP; for women: B = -5.104 + 0.051Age + SP). In males, the AUC was 0.914 (95% CI: 0.832–0.963), with a Youden's index value of 0.0.703 (Sen = 81.82%, Spe = 88.48%), and an optimal cut-off of 0.16. In females, the AUC was 0.848 (95% CI: 0.774–0.905), with a Youden's index value of 0.552 (Sen = 89.47%, Spe = 65.77%), and an optimal cut-off of 0.09. The predictive effect of the CVD model built by SP and age was better, and the predictive ability for CVD was better in Kazakhs men (Table 3 and Table 4).

Discussion

The standardized prevalence rate of MetS in Kazakhs was 28.61%, which was much higher than the national MetS prevalence rate of 16.5% [24]. Studies have shown that individuals with MetS are more likely to develop CVD [25,26], and a higher prevalence of MetS suggests that the risk of developing CVD in the Kazakh population may be higher.
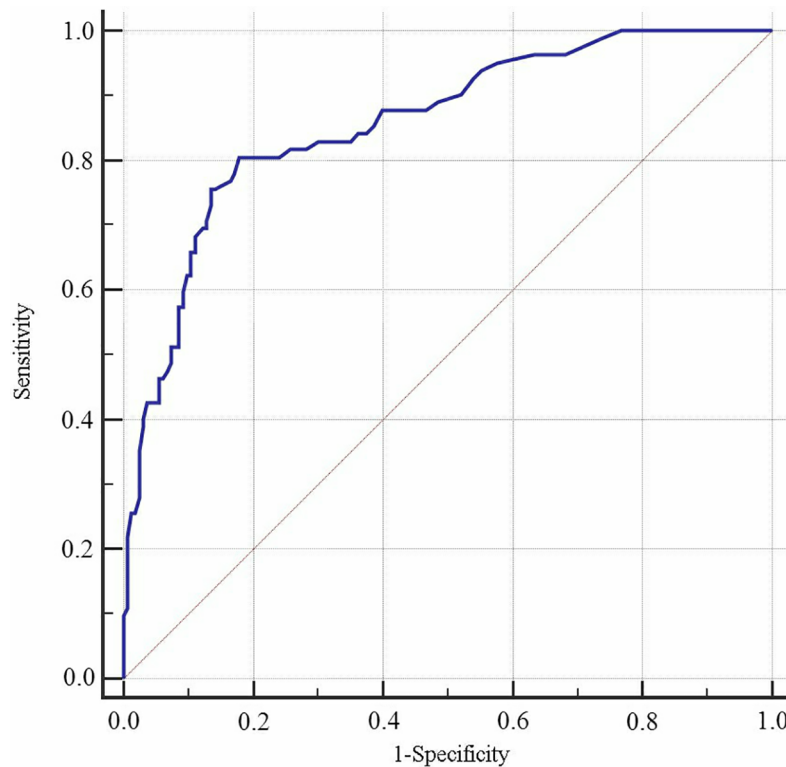
Fig 1. ROC curve for discrimination and prediction of CVD. This shows the discriminative curve under internal validation for male.

Based on the MetS multi-etiology hypothesis [27], this study used EFA to extract 7 principal components according to sex from18 risk factors that represent 7 potential factors for MetS occurrence in Kazakhs. Of the 7 potential factors, the first 3 factors for men and women were OF (WC & weight & BAI), HFF (TBIL & IBIL), and LF (HDL-C & APOA). Among these, obesity and dyslipidemia are classic components of MetS [28,29]. Body mass index (BMI) and WC are commonly used as indicators of obesity, and some studies have suggested that body fat content is a major risk factor for MetS, because the BAI is better than the BMI in estimating body fat content [30,31]. In this study, weight, WC, and BAI were used as the obesity factors, providing more comprehensive results than the W Zhang [32] study using BMI and WC as the obesity factors. The main characteristic of blood lipids in this study was the high prevalence of decreased levels of HDL-C [33]. Therefore, we extracted HDL-C and APOA as blood lipid factors in MetS, and the contribution of 2 lipid variables was greater than the contribution of the 4 variables reported in a previous study [34], indicating that the lipids extracted in this study are more accurate, and that these factors can better reflect the lipid accumulation characteristics in Kazakhs with MetS. In Kazakh MetS patients with HFF as the dominant factor, male sex as the second factor, and female sex as the first factor, the results are different from those in the Han Chinese population [34]. HFF may be an important factor in the assessment of MetS in Kazakhs. The contribution of EMT (ALT & AST & α-HBDH) was not high in our study, but was an important indicator for the diagnosis of non-alcoholic fatty liver disease. Non-alcoholic fatty liver disease has been shown to increase the risk of developing MetS, and the prevalence is higher in Kazakhs than in local Han Chinese [35, 36]. Therefore, ALT, AST and α-HBDH were selected as Kazakh MetS liver enzyme factors. In our study, the contribution of RMF (UA & CREA & BUN) was higher in men (9.00%) than in women (7.04%). In Wen-chao Zhang's
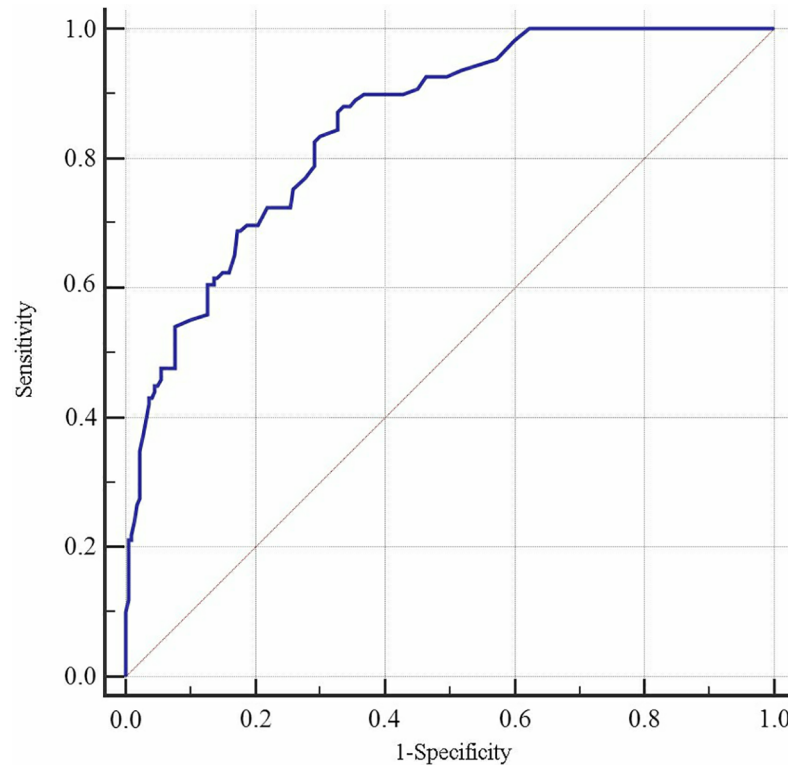
Fig 2. ROC curve for discrimination and prediction of CVD. This shows the discriminative curve under internal validation for female.

study [15], RMF was not extracted for men, but the female RMF contribution rate (6.81%) was similar to the rate in this study. It may be due to the reason that this was not included in CREA, leading to a weak RMF aggregation in men. Blood pressure and glucose metabolism in this study were male's 6th and 7th factors, and female's 4th and 7th factors, both contributing less than 10%. Consistent with the findings of A Ghosh [37], the contribution of blood pressure and glycemic factors was least. If a study subject was diabetic [38, 39], blood pressure and blood glucose were 1st and 2nd factors, suggesting that due to differences in study populations, the contribution of each factor to MetS was also different. Albumin is a common biochemical indicator and fluctuates in the human body due to disease. In this study, albumin was incorporated into male LF and female HFF, respectively. However, it was highly correlated with the routine physical examination indexes included in this study, and is retained for the accuracy of factor analysis.

In this study, a CVD predictive model was constructed using age and SP as predictors in the Kazakh population, with a predictive ability higher than that of the FRS (AUC approximately 0.80) [40]. This was also higher than the predictive ability achieved while using a Chinese evaluation method for 10-year risk of ischemic CVD (hereinafter referred to as the "Chinese evaluation method") in the optimal model (for men: AUC = 0.796; for women: AUC = 0.791) and the simple model (for men: AUC = 0.792; for women: AUC = 0.783) [41]. The Chinese evaluation method only used 8 risk factors as model predictors, which is the same as the number of predictive factors in our study. However, the potential factors in this study include obesity, blood pressure abnormalities, dyslipidemia, and glucose metabolism disorders. These cover the risk factors for CVD more comprehensively, and enhance the predictive ability of the model. Li Jiqing's findings showed that the male AUC in training samples was
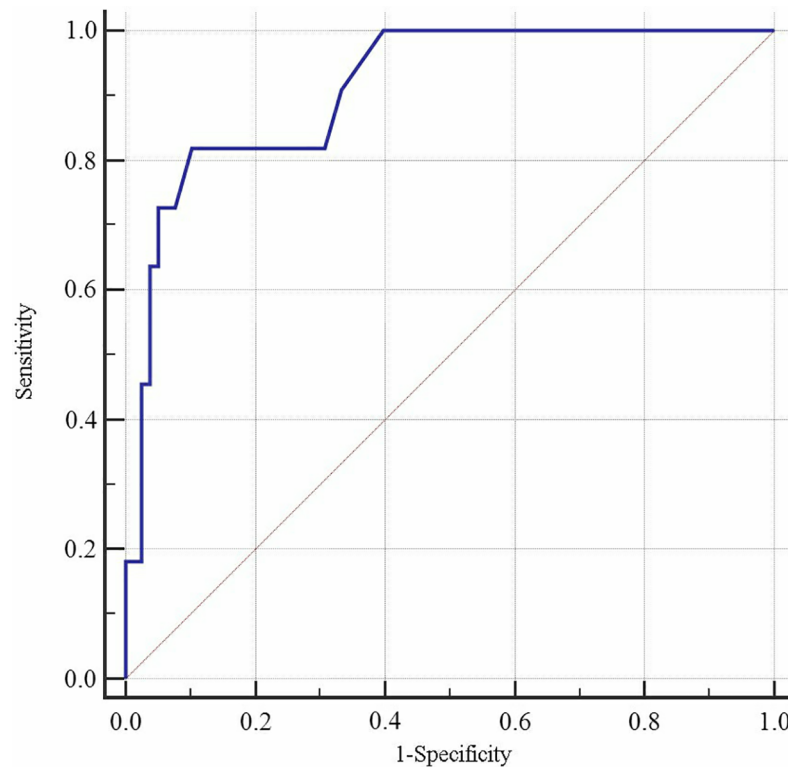
Fig 3. ROC curve for discrimination and prediction of CVD. This shows the discriminative curve under external validation for male.

lower than that in the current study (0.837 vs 0.857) [16]. However, the AUC in females was better than the AUC in this study (0.897 vs 0.852). Although the study still used the Chinese evaluation method, it also incorporated electrocardiographic information that was highly correlated with CVD. These data directly reflect CVD status, and can significantly improve the predictive and diagnostic capabilities of the model. Using MetS as a predictor of CVD, the current study only includes the classic components [42, 43], and overlooks the metabolic network of liver enzymes, renal function, bilirubin, and other components that affect CVD. However, Zhenxin Zhu's study [20] incorporated 8 factors of the MetS network into a CVD prediction model, with an AUC as high as 0.994 and 0.998 for men and women. We speculate that the present predictive model can fully cover the risk factors of CVD, which is key to improvement of the predictive ability. Therefore, predictors should cover both the risk factors and the incidence of CVD.

External verification is important for the assessment of the feasibility of the CVD model. The classic FRS model predicts a large predictive error for Hispanic, Asian, and elderly American populations, and significantly overestimates risk in a Han Chinese population [44], but has a better predictive ability for the Mongolian population [45]. Wang Fang's study of a Chinese health check-up group using the Chinese external validation method obtained an AUC of 0.717, which was significantly lower than the AUC in the optimal model and simple model [46]. It may be that because the outcome measured was carotid atherosclerotic plaque, the perdictive power was reduced. Using external verification to validate the established model, the AUC of men was 0.914 (95% CI: 0.832–0.963) and the AUC of women was 0.848 (95% CI: 0.774–0.905), both of which were greater than or equal to 0.85. This shows that the CVD model can be applied to local Kazakhs, especially for male Kazakh CVD prediction. Based on
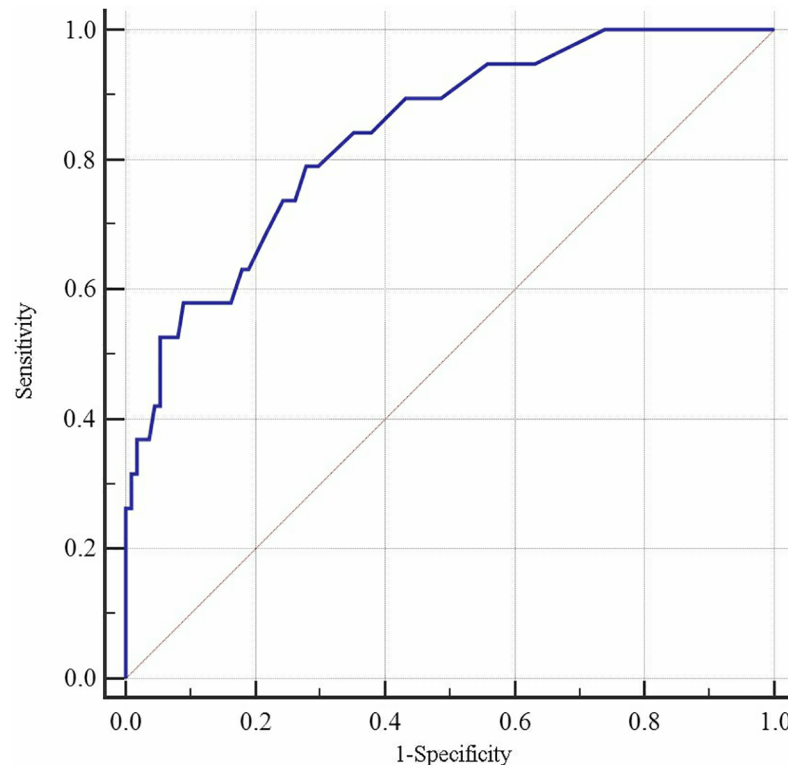
Fig 4. ROC curve for discrimination and prediction of CVD. This shows the discriminative curve under external validation for female.

age and SP, this study can be used to predict the incidence of CVD in the local Kazakh population, which is conducive to early screening in primary health institutions, and can thus help reduce the overall incidence of CVD by controlling risk factors.

Our study had the following advantages. First, the study is based on the characteristics of local Kazakh disease. The prevalence rates of MetS, hypertension, and obesity in Kazakhs are significantly obviously higher than those in other ethnic groups residing in the same area. MetS was used to construct a CVD prediction model for the Kazakh population, as similar studies have not been reported. Second, the results of this study are based on indicators used in routine physical examinations. The indicators are all derived from national health checkup projects supported by the Chinese government. These indicators are easy to determine and are cost effective. Third, this study used EFA to extract 7 potential factors from 18 routine physical examination indicators to form a SP, and to a greater extent to strengthen the predictive ability of the model. This was not only a simple summation of factors, but also showed a correlation between CVD prediction results and potential factors. Finally, the model underwent internal and external validations to illustrate the applicability. Therefore, this model can be applied to local Kazakh residents for early prevention of CVD and for screening of high-risk individuals.

This study had some limitations. First, routine physical examination indicators cannot cover all risk factors for CVD, resulting in limited predictive ability for the model. Second, the sample size is not big enough. Since the Kazakh lifestyle is mainly nomadic, data were not collected for CVD patients who did not go to the hospital or those whose hospitalization records were missing. Therefore, this study may underestimate the cumulative incidence of CVD. Subsequent research can supplement follow-up information, increase the number of routine

Table 3. Results of logistic regression discrimination for internal and external validation in males and females.

Variables	Model 1 B	P-value	OR(95%)	Model 2 B	P-value	OR(95%)
Male						
Constant	0.733	0.442	2.081	-7.550	P<0.001	0.001
Age	-0.039	0.038	0.962(0.928–0.998)	0.079	0.019	1.082(1.013–1.156)
OF	1.561	P<0.001	4.765(2.839–7.996)	0.683	0.101	1.980(0.875–4.479)
HFF	-0.244	0.168	0.784(0.554–1.108)	0.555	0.047	1.742(1.007–3.014)
LF	-0.597	0.004	0.551(0.367–0.826)	-0.576	0.268	0.562(0.203–1.557)
RMF	0.259	0.190	1.295(0.880–1.906)	0.167	0.740	1.182(0.441–3.170)
EMF	-0.372	0.101	0.690(0.442–1.075)	0.363	0.415	1.438(0.601–3.442)
BPF	0.636	0.003	1.888(1.247–2.860)	0.549	0.172	1.732(0.788–3.808)
GMF	0.624	0.002	1.867(1.247–2.795)	-0.051	0.912	0.950(0.383–2.355)
Female						
Constant	0.077	0.916	1.08	-5.104	0.002	0.006
Age	-0.016	0.260	0.984(0.956–1.012)	0.051	0.059	1.053(0.998–1.111)
HFF	-0.511	0.003	0.600(0.427–0.843)	0.049	0.895	1.051(0.504–2.191)
OF	1.168	P<0.001	3.217(2.248–4.602)	0.972	0.008	2.642(1.291–5.409)
LF	-0.61	P<0.001	0.543(0.387–0.763)	-0.220	0.539	0.802(0.397–1.621)
BPF	0.693	P<0.001	1.999(1.399–2.858)	0.674	0.036	1.961(1.044–3.685)
EMF	-0.279	0.103	0.756(0.541–1.058)	0.125	0.680	1.133(0.626–2.052)
RMF	0.472	0.009	1.603(1.124–2.287)	-0.325	0.427	0.723(0.325–1.609)
GMF	0.892	P<0.001	2.440(1.610–3.698)	0.583	0.186	1.791(0.755–4.246)

Note: Model 1, Internal validation. For male: B = 0.733–0.039Age + SP, SP = 1.561OF—0.244HFF—0.597LF + 0.259RMF—0.372EMF + 0.636BPF + 0.624GMF; for female: B = 0.077–0.016Age+SP, SP = -0.511HFF + 1.168OF—0.610LF + 0.693BPF—0.279EMF +0.472RMF + 0.892GMF. **Model 2**, External validation. For male: B = -7.55 + 0.079Age + SP, SP = 0.101OF + 0.555HFF—0.576LF + 0.167RMF + 0.363EMF + 0.549BPF—0.051GMF; for female: B = -5.104+0.051Age+SP, SP = 0.049HFF + 0.972OF—0.220LF + 0.674BPF + 0.125EMF—0.325RMF + 0.583GMF.

https://doi.org/10.1371/journal.pone.0202665.t003

Table 4. Results of ROC curves for internal and external validation in males and females.

	Group	Cut-off	Sen(%)	Spe(%)	Youden's index	AUC(95%CI)	P-value
Model 1	Male	0.30	80.49	81.71	0.622	0.857(0.807–0.898)	P<0.001
	Female	0.23	88.07	66.36	0.544	0.852(0.809–0.889)	P<0.001
Model 2	Male	0.16	81.82	88.46	0.703	0.914(0.832–0.963)	P<0.001
	Female	0.09	89.47	65.77	0.552	0.848(0.774–0.905)	P<0.001

Note: Model 1, Internal validation; Model 2, External validation. Abbreviation: Sen-sensitivity; Spe-specificity.

https://doi.org/10.1371/journal.pone.0202665.t004

physical examination indicators, extract potential factors that can cover more comprehensive information, and improve the predictive ability of the model.

Supporting information

S1 Table. Distribution of age and the eighteen biomarkers between male and female groups. (DOCX)

S2 Table. Correlation matrix of eighteen biomarkers. (DOC)

S3 Table. The standardized scoring coefficients of each factor for males and females.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Shuxia Guo.

Data curation: Lei Mao, Kui Wang.

Formal analysis: Lei Mao, Kui Wang, Wenwen Yang, Shuxia Guo.

Investigation: Lei Mao, Jia He, Heng Guo, Xianghui Zhang, Wenwen Yang, Jingyu Zhang, Shugang Li, Yunhua Hu, Lati Mu, Yizhong Yan, Jiaolong Ma, Yusong Ding, Mei Zhang, Jiaming Liu, Rulin Ma, Shuxia Guo.

Project administration: Shugang Li.

Writing – original draft: Lei Mao, Jia He, Shuxia Guo.

Writing – review & editing: Lei Mao, Xiang Gao, Rulin Ma, Shuxia Guo.

References

1. Weiwei C, Runlin G, Lisheng L, Manlu Z, Wen W, Yongjun W, et al. Report on cardiovascular diseases in China (2016). Chinese Circulation Journal. 2017; 32(6):521–530.

2. Isohwg WHO. Definition, diagnosis and classification of diabetes mellitus and its complications: report of a World Health Organisation Collaborations. Diabet Med. 1999; 15(7):539–553.

3. Alberti G, Zimmet P, Shaw J, Grundy SM. The IDF consensus worldwide definition of the metabolic syndrome. Hop Tests As Predictors of Dynamic Knee Stability. 2006.

4. Xiao Y, Chen J, Chen W. The Relationship between Hyperuricemia and Metabolic Syndrome. Journal of Baotou Medical College. 2011; 27(3):47–49.

5. Tan CE, Ma S, Wai D, Chew SK, Tai ES. Can we apply the National Cholesterol Education Program Adult Treatment Panel definition of the metabolic syndrome to Asians? Diabetes Care. 2004; 27 (5):1182. PMID: 15111542

6. Alberti KG, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International A. Circulation. 2009; 120(16):1640. https://doi.org/10.1161/CIRCULATIONAHA.109.192644 PMID: 19805654

7. Chen S, Guo X, Yu S, Zhou Y, Li Z, Sun Y. Metabolic Syndrome and Serum Liver Enzymes in the General Chinese Population. International Journal of Environmental Research & Public Health. 2016; 13 (2):223. https://doi.org/10.3390/ijerph14020156

8. Hou XH, Zhu YX, Lu HJ, Chen HF, Li Q, Jiang S, et al. Non-alcoholic fatty liver disease's prevalence and impact on alanine aminotransferase associated with metabolic syndrome in the Chinese. J Gastroenterol Hepatol. 2011; 26(4):722–730. https://doi.org/10.1111/j.1440-1746.2010.06509.x PMID: 21418302

9. Choi SH, Yun KE, Choi HJ. Relationships between serum total bilirubin levels and metabolic syndrome in Korean adults. Nutrition Metabolism & Cardiovascular Diseases. 2013; 23(1):31.

10. Lin LY, Kuo HK, Hwang JJ, Lai LP, Chiang FT, Tseng CD, et al. Serum bilirubin is inversely associated with insulin resistance and metabolic syndrome among children and adolescents. Atherosclerosis. 2009; 203(2):563–568. https://doi.org/10.1016/j.atherosclerosis.2008.07.021 PMID: 18775539

11. Nejatinamini S, Ataiejafari A, Qorbani M, Nikoohemat S, Kelishadi R, Asayesh H, et al. Association between serum uric acid level and metabolic syndrome components. Journal of Diabetes & Metabolic Disorders,14,1(2015-09-14). 2015; 14(1):1–7.

12. Liu X, Zhang H, Liang J. Blood urea nitrogen is elevated in patients with non-alcoholic fatty liver disease. Hepato-gastroenterology. 2013; 60(122):343. PMID: 23858556

13. Jing Yuling, Hong, Ralph, D', Agostino, et al. Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the chinese multi-provincial cohort study.The Journal of the American Medical Association(Chinese Edition). 2006;(2):67–74.

14. Chen JH, Wu HY, He KL, He Y, Qin YH. Establishment of the prediction model for ischemic cardiovascular disease of elderly male population under current health care program. Chinese Journal of Epidemiology. 2010; 31(10):1166. PMID: 21162823

15. Zhang W. Metabolic Syndrome prediction models and tools based on longitudinal Health Management Cohort Data: Shandong University; 2016.

16. Jiqing LI, Huanzong Z, Binghong S, Lichun Z, Xiangyi L, Yafei C, et al. Risk prediction model of cardiovascular disease based on health management cohort. Shandong University (Health Science). 2017; 55(6):56–60.

17. Tunstall-Pedoe H. The Dundee coronary risk-disk for management of change in risk factors. Bmj. 1991; 303(6805):744–747. PMID: 1932933

18. Voss R, Cullen P, Schulte H, Assmann G. Prediction of risk of coronary events in middle-aged men in the Prospective Cardiovascular Munster Study (PROCAM) using neural networks. International Journal of Epidemiology. 2002; 31(6):1253–1262. PMID: 12540731

19. Hippisleycox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. Bmj. 2008; 336 (7659):1475–1482. https://doi.org/10.1136/bmj.39609.449676.25 PMID: 18573856

20. Zhu Z, Liu Y, Zhang C, Yuan Z, Zhang Q, Tang F, et al. Identification of cardiovascular risk components in urban Chinese with metabolic syndrome and application to coronary heart disease prediction: a longitudinal study. Plos One. 2013; 8(12):e84204. https://doi.org/10.1371/journal.pone.0084204 PMID: 24358344

21. Xianghui Z, Yizhong Y, Jia H, Jiaolong M, Hongrui P, Rulin M, et al. Evaluation of screening indices for metabolic syndrome in adult Kazakh in Xinjiang. Chinese Journal of Hypertension. 2017(1):55–60.

22. Guo SX, Zhang JY, Zhang YH, Shang-Zhi XU, Guo H, Niu Q, et al. Epidemiological Characteristics and Analysis on Hypertension of Kazakstans and Hans in Xinjiang. Journal of Shihezi University. 2008.

23. He J, Guo H, Ding YS, Liu JM, Zhang M, Ma RL, et al. Epidemiological study on overweight and obesity among rural adult residents in Hazakh,Uygur and Han populations in Xinjiang. Chinese Journal of Epidemiology. 2013; 34(34):1164–1168.

24. Dongfeng G, Reynolds K, Wenjie Y, Shufeng C, Xigui W, Xiufang D, et al. The prevalence of metabolic syndrome in the general adult population aged 35–74 years in China. Chinese Journal of Diabetes Mellitus. 2005; 13(3):181–186.

25. Novo S, Peritore A, Guarneri FP, Corrado E, Macaione F, Evola S, et al. Metabolic syndrome (MetS) predicts cardio and cerebrovascular events in a twenty years follow-up. A prospective study. Atherosclerosis. 2012; 223(2):468–472. https://doi.org/10.1016/j.atherosclerosis.2012.05.018 PMID: 22704563

26. Dada AS, Ajayi DD, Areo PO, Raimi TH, Emmanuel EE, Odu OO, et al. Metabolic Syndrome and Framingham Risk Score: Observation from Screening of Low-Income Semi-Urban African Women. Medicines. 2016; 3(2):15.

27. R K, J B, E F, M S. The metabolic syndrome: time for a critical appraisal. Joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. Diabetologia. 2005; 48(9):1684–1699. https://doi.org/10.1007/s00125-005-1876-2 PMID: 16079964

28. Fan HQ, Guo XR. Obesity and Metabolic Syndromes. Medical Recapitulate. 2007; 13(1):71–73.

29. Chen MJ, Xiang-Ping LI. Characteristics of and Therapy for Dyslipidemia in Metabolic Syndrome. Advances in Cardiovascular Diseases. 2011; 32(1):104–107.

30. Yan SJ, Li YM, Shen XM, Yang LY, Zhang SJ, Yi RH, et al. Use body fat determination instead of simple body composition parameters evaluate the risk of metabolic syndrome in Fuzhou adults. National Medical Journal of China|Natl Med J China. 2010; 90(24):1678–1681. PMID: 20979876

31. Jie L, Lingling Q, Huating L, Zaixiang T, Cheng H, Chen W, et al. Application of body adiposity index in the Chinese people in Shanghai area. Chinese Journal of Diabetes Mellitus. 2016; 8(5):278–282.

32. Zhang W, Chen Q, Yuan Z, Liu J, Du Z, Tang F, et al. A routine biomarker-based risk prediction model for metabolic syndrome in urban Han Chinese population. BMC Public Health. 2015; 15(1):1–9.

33. Ru-Lin MA, Guo SX, Yan LI. Prevalence of dyslipidemia and its influencing factors in Kazakh adults. Chinese Journal of Public Health. 2012; 28(8):1009–1013.

34. Cuifang Y. Cohort study on risk prediction model of metabolic syndrome based on routine physical examination biomarker by exploratory factor analysis: Dalian Medical University; 2017.

35. Adams LA, Waters OR, Knuiman MW, Elliott RR, Olynyk JK. NAFLD as a risk factor for the development of diabetes and the metabolic syndrome: an eleven-year follow-up study. American Journal of Gastroenterology. 2009; 104(4):861–867. https://doi.org/10.1038/ajg.2009.67 PMID: 19293782

36. Zhang W, Sha H, Qin J, Nina YE, Huo X, Gastroenterology DO. Epidemiological Survey of Nonalcoholic Fatty Liver Disease in Yili Area. Clinical Medicine & Engineering. 2013; 20(9):1169–1170.

37. Ghosh A. Factor analysis of metabolic syndrome among the middle-aged Bengalee Hindu men of Calcutta, India. Diabetes/metabolism Research & Reviews. 2005; 21(1):58–64.

38. Sun F, Tao QS. The analysis of the components of metabolic syndrome among diabetic patients by use of factor analysis and factor scores. Chinese Journal of Diabetes. 2010; 18(11):818–820.

39. Gong M. Factor Analysis of Components of the Metabolic Syndrome. Guide of China Medicine. 2011; 09(2):14–17.

40. De LIB, Potter JF, Poulter NR, Robins MM, Skinner J. Performance of the ASSIGN cardiovascular disease risk score on a UK cohort of patients from general practice. Heart. 2011; 97(6):491–499. https://doi.org/10.1136/hrt.2010.203364 PMID: 21097820

41. Wu Y, Liu X, Li X, Li Y, Zhao L, Chen Z, et al. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. Circulation. 2006; 114(21):2217–2225. https://doi.org/10.1161/CIRCULATIONAHA.105.607499 PMID: 17088464

42. Li-Li YU, Lu XU, Shao JH. Comparative Study on Prediction of Cardiovascular Disease Risk by Metabolic Syndrome and Component Clusters. Chinese General Practice. 2012; 15(3):241–244.

43. Sun HJ, Jang Y, Dong JO, Oh BH, Sang HL, Park SW, et al. A coronary heart disease prediction model: the Korean Heart Study. Bmj Open. 2014; 4(5):e005025. https://doi.org/10.1136/bmjopen-2014-005025 PMID: 24848088

44. Chen JH, He Y, Qin YH. A review of cardiovascular disease prediction model at home and abroad. Chinese Journal of Epidemiology|Chin J Epidemiol. 2010; 31(11):1311–1314. PMID: 21176700

45. XIaohong Z. Study on the risk factors of cardio cerebral vascular disease and the applicability of Framingham score in Mongolian population: Soochow University; 2014.

46. Fang W, Lirong L, Liping Y, Guangzhen Z. Value of Chinese Evaluation Method Versus Pooled Cohort Equations for Assessing the 10-year Risk of Ischemic Cardiovascular Disease in Population Undergoing Health Examination. Chinese General Practice. 2017; 20(28):3498–3503.